EPA Comments Appendix E: Round 2 Comprehensive Site Characterization Summary and Data Gaps Report
June 13, 2008

Introduction:

EPA has previously provided comments on various sections the Round 2 Comprehensive Site Characterization Summary and Data Gaps Report (Round 2 Report). These comments on Appendix E of the Round 2 Report represent EPA's final comments on the Round 2 Report.

Appendix E presented information on the development of biota-sediment accumulation factors (BSAFs) and the site specific food web model for the Portland Harbor Superfund Site. These evaluations are considered critical to the development of cleanup goals for bioaccumulative chemicals such as polychlorinated biphenyls (PCBs), organochlorine pesticides (e.g., 4,4'-DDT) and polychlorinated dibenzo-p-dioxins and furans (PCDD/PCDF).

General Comments:

Model Parameterization:

The Portland Harbor Foodweb model relied on a series of modifications to the Arnot and Gobas model. In particular, the model included the optimization of 21 internal parameters. It is unclear whether these modifications were appropriate in all cases. In general, EPA does not support modification of the internal parameters of the Arnot and Gobas model. Modification of these parameters suggest that the Portland Harbor data is of better quality than those used by Arnot and Gobas to build and validate their model; this may not be the case. Given the complexities of the site, it is unclear whether the site specific data are in fact superior and should be used as the basis for parameterization of the model. Instead of fine tuning values to get agreement, it may be more appropriate to evaluate the uncalibrated model to test model performance.

Use of Surface Weighted Average Concentrations:

The Portland Harbor Foodweb model relied on the use of surface weighted average concentrations (SWACs) to develop average sediment concentrations. The use of a site wide SWAC value assumes that the organism uses the entire site. Further, the SWAC was computed assuming equal chemical bioavailability across the ~~site~~. Based on the laboratory worm and clam bioaccumulation studies, an assumption of equal chemical bioavailability may not be appropriate. Because model is constructed using a single water and sediment input concentrations, some type of averaging will be required. One possibility that should be considered is to drive the food web model with residues in the worm and clam data rather than the residues in the sediment in order to consider chemical bioavailability

Evaluation of Model Performance:

Arnot and Gobas recommend evaluating the ratio of predicted and observed concentrations to test model performance. Currently, the Portland Harbor food web model does not use an

estimate of observed vs. predicted values to provide an estimate of uncertainty because of its over calibration. The data may be used to determine how uncertain the model is. This is different than model uncertainty analysis, which should be avoided because many parameters are correlated. It is therefore difficult to apply Monte Carlo simulations of uncertainty. The only parameters that are NOT correlated and a true uncertainty analysis could be performed would be lipid content, temperature, and K_{ow}.

Model predictions for TEQ PCBs are no better than between 5x and 10x of the empirical data. Because the model appears to work well for other non-polar organic chemicals, investigation into the cause of this poor performance is recommended. For example, a review of the TEQ data to ensure that data quality issues are not resulting in poor TEQ performance may be warranted.

Modeling of Chemical Mixtures:

For PCB modeling, Gobas has found that you should be able to get within a factor of 2 to 2.5 – these factors shouldn't be any higher. EPA cautions against modeling total PCBs and recommends focusing on individual congeners only. The calibrated food web model was used to make predictions for chemical groups, e.g., total DDTs, total PCBs, PCB TEQs (birds), PCB TEQ (mammals), Dioxin TEQ (birds), and Dioxin TEQ (mammals). Predictions for mixtures are predicated upon have the correct weighted average K_{ow} of the mixture. If the composition of the mixture is not the same across the harbor, predictions will be inaccurate. For example, the SWAC for the surface sediment averages the total from each surface sample and if the compositions vary widely among surface samples, selecting an appropriate K_{ow} will be difficult.

Chemical Bioavailability:

A review of the data suggests that chemical bioavailability changes with increasing concentration of the chemical in the sediment on an organic carbon basis. For example, BSAFs decline with increasing concentration in the sediment. Because the laboratory testing confirms the field observation, it appears that the observed trend is real. There are a number reasons why chemical bioavailability could be changing. First, the presence of black carbon, and second, the presence of oil/grease like phase. Both of these phases have much higher sorptive capacity for the organic contaminants of interest, and their presence in a harbor would not be unexpected.

Because the spreadsheets of data don't contain the organic carbon content of the sediments, it is not possible to determine if the BSAFs decrease with increasing organic carbon content. If so, this observation would point one in the direction of having a black carbon and/or oil/grease phases present.

The observation of declining bioavailability with increasing concentrations in the sediment clearly has implications for the development of sediment PRGs and the accuracy of long term tissue predictions through the hybrid model.

Quality of Field Data:

Commented [MM2]: Is this true? This is Jennifer's comment. However, it seems to me that the SPAF presented in the report evaluates model performance.

Commented [jp3R2]: My point was that the current model does not evaluate predicted versus observed with an uncalibrated Gobas model. Instead, the model was calibrated to reduce this variability. The calibration in essence hides model incongruity and uncertainty that may not be appropriate – i.e. should we be changing all those parameters? Do we have more info than the Gobas effort to change values with an appropriate level of confidence?

The development of BSAFs and the Portland Harbor food web are dependent upon the quality of the field data. Overall, EPA believes that the Portland Harbor RI/FS has generated high quality data. However, a review of the some of the data during the review of the food web model has identified some anomalies. For example, the PCDD and PCDF data seem out of line with the PCBs in their bioaccumulation behavior. In addition, sediment sample "LWG0107R006SDS015C00" (located off-shore of Arkema) has usually high PCDF values relative to other samples and to its own PCDD levels; i.e., 2,3,7,8-TCDF is 14,000 pg/g while 2,3,7,8-TCDD is 5.4 pg/g, and other samples are 1,000-10,000 fold lower for 2,3,7,8-TCDF. EPA recommends a review of these results to ensure that the data is accurate.

Next Steps and Recommendations:

Food web models generally require a large number of input parameters. For the Arnot and Gobas model, 70 input parameters are required. As noted above, for the Portland Harbor site, 21 internal parameters were optimized resulting a total of 91 parameters being optimized. Given the limited number of average values used in the Monte Carlo process, the model appears to be highly calibrated to the field data used in the calibration process and what ever deficiencies exist in these data are captured in the calibration process. EPA recommends evaluating the model against the recently collected Round 3B tissue data to develop an understanding the accuracy of the predictions from Portland Harbor model. It is unclear whether the addition of the additional 21 parameters improves or reduces model performance. If the model does not perform well with the new data, EPA recommends reverting back to the uncalibrated Arnot and Gobas model.

In general, EPA recommends modeling each chemical individually, and then, combining them after making the predictions. This avoids the messy averaging done prior to the make the food web predictions and would also eliminate the problems with differing compositions of the mixture across the site. The predictive ability of the model might improve using this approach for mixtures.

Specific Comments:

Section 2.0 – Biota-Sediment Relationship Development:

Based on information presented in Appendix E and errata information presented on the Portland Harbor Portal, it is unclear whether a regression approach or an averaging approach was employed for the development of biota-sediment accumulation functions (BSAFs). Further, the report does not present the estimated BSAFs for benthic organisms (e.g., field and laboratory clams, crayfish, sculpin and laboratory worms). The approach for developing BSAFs and the estimated BSAFs should be clearly presented in the draft Remedial Investigation (RI) Report.

In general, EPA recommends the use of the paired approach rather than the regression approach. A quick review of the estimated BSAFs for clam tissue indicates a wide range in sample specific BSAFs. For example, calculated BSAFs for Aldrin in field clams range from 0.325 and 15. EPA believes that BSAFs should be evaluated to identify trends in the underlying ecological conditions at the sampling location (e.g., food web structure, sediment/water column relationships, chemical bioavailability, and diets/trophic levels of the organisms. BSAFs should

Commented [MM6]: I would like to be clear about next steps. On one hand, both Larry and Jennifer have recommended not calibrating the model. On the other hand, Larry seems to suggest a validation step using the Round 3B data. I have split the difference somewhat with my recommendation. Does this make sense?

Commented [jp7R6]: Evaluating Round 2 food web model with the Round 3 data may help show where the model is insufficient. However, more importantly the model should be "uncalibrated" and the all observed values (Round 1 and 3) compared to the predicted values. This will give us the best evaluation of model performance and identify where in the harbor tissue is under or over predicted.

be plotted spatially and plots of BSAF against lipid normalized tissue concentrations should be presented in addition to the plots of BSAF against organic carbon normalized sediment concentrations.

The plots of BSAF against concentrations of the chemical in the sediment on an organic carbon basis reveals a trend of decreasing BSAF with increase concentration in the sediment. For the laboratory derived BSAFs for worms and clams, the BSAFs are generally below 10 for low concentrations in sediment. Field clams BSAF have the same behavior. This generally argues against the use of a regression analysis to develop BSAFs and has implications for the performance of the food web model.

Although the regression analyses appear to have been properly performed, the equations in BSAF_eqns_shellfish_052907.xls don't line up exactly with data in FieldClams_BSAFs.xls. More significantly, some of the regressions have a low predictive power, (e.g., Aldrin has a r^2 of 16% for shellfish). This further suggests that a averaging approach should be utilized.

EPA's independent review of the BSAFs across the site suggest that BSAFs may vary spatially across the site. Further evaluation of BSAF trends may prove useful in developing location specific BSAFs or developing additional normalization techniques that may reduce the uncertainty in BSAF estimates.

Section 2.1 – Benthic Invertebrate/Sculpin BSAFs

BSAFs were developed for clams, crayfish and sculpin based on field collected data. In addition, BSAFs were developed for clams and worms based on laboratory bioaccumulation tests. Data was evaluated to determine whether an average paired BSAF or a regression line BSAF should be developed. Based on information presented in the report and the errata section of the Portland Harbor Portal, it appears that a regression approach was used to develop BSAFs for clam tissue and that this relationship was applied to all shellfish. However, it does not appear that the supporting information was included (e.g., strength of regression, statistical testing results, etc. Furthermore, no table is included in Appendix E that summarizes the BSAFs for field collected clams, crayfish and sculpin and laboratory testing of clams and worms. This information should be presented.

Section 2.2 – Demersal/Pelagic Fish BSAFs:

EPA does not believe it is appropriate to develop BSAFs for the majority of fish species collected at the Portland Harbor site. This recommendation is primarily based on the fact that the majority of fish species tested have limited contact with contaminated sediment, the confounding factors of dietary and water column exposure, and uncertainty about where the fish may have come into contact with contaminated sediment. As a result, the only fish species (in addition to sculpin) for which BSAFs may reasonably be estimated are carp due to their propensity to accumulate chemicals, high lipid content and frequent contact with sediment during feeding.

Commented [MM9]: I added this comment based on a statement Burt has made in the past. However, we will be developing tissue-residue TRVs for more chemicals than are modeled in the FWM and for more fish species than carp. How do we develop PRGs for these chemicals and fish if we do not apply a BSAF. Maybe it is better to develop BSAFs for species other than carp and just acknowledge the uncertainty.

Section 3.2, Model Development and Methodology:

This section of the report identifies two key underlying assumptions of the Arnot and Gobas food web model: 1) The system is in steady state and 2) the flux between different media are governed by fugacity relationships. As a result of these assumptions, the food web model is expected to work best for chemicals that are subject to relatively fast exchange kinetic (e.g. lower K_{ow} substances) and small organisms because steady state will be achieved rapidly in these situations. The corollary is that the model should be used with caution in situation where the exchange kinetics are very slow (e.g. chemicals that are metabolized slowly or chemicals with high K_{ow} in large lipid rich organisms) because steady state takes a long time to achieve.

Section 3.2.5.1, Selection of Model Parameter Values and Distributions Used for Calibration:

EPA and the LWG agreed to consider a range of certain model parameters (dietary matrix, average water temperature, average body weights, average lipid content, K_{ow} and average sediment and water concentrations) in order to support a sensitivity analysis. However, the food web model varies many other parameters beyond this list such as moisture content, dietary absorption efficiency of lipid, dietary absorption efficiency of NLOM, dietary efficiency of water, and fraction of porewater ventilated. In addition to the input parameters a whole series of parameters, internal to the Arnot and Gobas model, were adjusted by using the Monte Carlo analysis. These additional internal parameters are: 1) Dietary absorption efficiencies for lipid (eL) and NLOM (eN); 2) phytoplankton uptake constants A and B; and 3) NLOM proportionality constant BETA. Dietary absorption efficiencies were optimized on a species specific basis. These parameters are presented in Tables 3-5 to 3-9.

Examination of Tables 3-5, 3-6, and 3-7 reveals that the calibrated values for environmental, general biological, and species-specific parameters are largely unchanged from those used by Arnot & Gobas except for: 1) Crayfish dietary absorption efficiency of lipid and NLOM; 2) clams dietary absorption efficiency of lipid; and 3) phytoplankton uptake constants A and B.

From Tables 3-5, 3-6, 3-7 and 3-8, there are: Four environmental parameters, three general biological parameters of which three are internal parameters; fifty-two species-specific biological parameters of which eighteen are internal parameters; and thirty-two species-specific dietary parameters. This results in ninety-one total parameters (of which 21 are Arnot & Gobas internal parameters) that are optimized in the Monte Carlo procedure. There are 10 different species in the food web and only 5 of the species (clams, carp, crayfish, sculpin, and smallmouth bass) have values for PCBs 17, 170 and 206. These five values are used in calibrating the model, i.e., providing the "calibrated" values reported in Tables 3-5, 3-6, 3-7 and 3-8. It is unclear how the "calibrated" values for the largescale sucker and northern pikeminnow are derived using PCBs 17, 170, and 206 when these species have no data for these congeners (see Appendix E2, Table 1). For the lower trophic levels, i.e., phytoplankton, zooplankton, and benthic invertebrates, their calibrated values can be found based upon the their predator residues.

In general, Arnot and Gobas do not recommend calibrating the food web model as long as the model is used with its application domain. However, unless sufficient information (e.g., site specific empirical data) is available to represent these parameters as a robust distribution or unless we have conducted additional site specific studies to justify a change to the value, EPA

recommends against modifying the values to get better agreement between the model and tissue. This is especially true given the inherent uncertainties of the data set and the dynamic nature of the Willamette River.

In light of the additional parameterization of the model, it is important to check whether the model calculations are consistent with available empirical data such that confidence in the model is gained. This has the same objective of calibration, but it is more of a hypothesis testing scenario, where the model outcomes are compared with independent data (data not used in the construction of the model). In this case, the preferred model outcome is the chemical concentration in a particular species, and involves the comparison of observed and predicted concentrations in different aquatic species.

Table 3-7 presents "calibrated values" for weight, lipid content, etc. for each species. However, instead of taking this approach, a rigorous evaluation of how well the model outcomes match tissue observations in Portland Harbor should be performed. In order to evaluate how well the model outcomes match the observations, Gobas uses the mean model bias MB to express the central tendency of the model. MB is the geometric mean (assuming a log-normal distribution) of the ratio of predicted and observed concentrations. Equation 3 on page 13 presents a similar equation. However, this equation does not appear to use a log-normal distribution, and does not appear to use information from a number of observations from each species (it looks like one average is used).

From a modeling perspective, the model codes are those of Arnot and Gobas and the only differences between the Portland Harbor model and the Arnot & Gobas model are the values for the 21 internal parameters discussed above. EPA would like to point out that Arnot and Gobas went to great lengths in assembling data that were of high quality across a number of food webs and these data have coverage of all species in these food webs. Subsequently, they used these data to validate their model. The effort here assumes that the Portland Harbor data are of higher quality or otherwise better than that used by Arnot and Gobas. However, given the complex nature of the Portland Harbor site and issues related to bioavailability, it is unclear whether the Portland Harbor data are of higher quality. In addition, it is unclear how the 21 internal parameters were selected for optimization; why not all the parameters in the model such as equations for growth and respiration? The modeling approach presented in Appendix E suggests that by optimizing these 21 internal parameters, the Portland Harbor model will have better predictive power than the Arnot & Gobas model. Mathematically with the 5 average values, the Portland Harbor model does provide a good fit, i.e., SPAFs are small. However it is likely that the Arnot and Gobas model (with Arnot and Gobas parameterization) would provide similar fits.

Although EPA questions the appropriateness of the calibration step, it does not appear that the execution of the Monte Carlo technique was performed improperly. The SPAFs, and sensitivity and uncertainty analyses demonstrates that Portland Harbor model does fit the field data from the site reasonably well. These analyses don't speak or address the issue of the quality of field data. The quality of the predictions from the Portland Harbor model are predicated upon the quality of the field data and if quality of the data is low, the model will provide poor predictions, in essence, garbage in – garbage out. To address the issue of the quality of the field data, additional

data sets from time periods totally independent of the calibration data are required. With these data, predictions based upon the conditions (concentrations of chemical in sediment and water) can be made and then, compared to independent data. These would provide a much better indication of the predictive power of the Portland Harbor model.

Given the number of parameters being optimized with the Monte Carlo technique and the relatively limited amounts of the field data, the model seems over parameterized. Without the 21 internal parameters, there are 70 input values being optimized with the Monte Carlo technique. Of the 70 inputs values, 32 are diet, 9 are weight, 10 are lipid content, and 10 are moisture content, and all of these values are required input values for the model. Unfortunately, this is the nature of food web model.

Commented [MM13]: There are a lot of comments from Jennifer and Larry regarding the parameterization step. These need to be reviewed for flow and consistency. I took my best shot.

Section 3.2.5.1.1 – Chemical Concentrations in Water:

Understanding the relative contribution of sediment contaminants and water contaminants to fish tissue levels is critical from the standpoint of PRG development. The appropriate place to perform this analysis is with the empirical data prior to PRG development. There are multiple possibilities in achieving acceptable fish tissue concentrations with the back calculation. The most straight forward way to address this question is with the use of empirical data. This will entail a discussion on how water is handled in the model. EPA believes that use of a total water concentration~~s~~ overestimates the tendency of contaminant in water to partition to fish tissue. This is especially true where a lot of TSS or particulate organic matter is present in the water. The chemical will be bound up on particulates and not very available to the fish. As a result, dissolved measurements from the field should be utilized to ensure that this overestimate does not occur. In this version of the food web model, a generic equation was applied to go from total to dissolved concentrations. EPA believes that the use of empirical surface water data is preferable to a calculated estimate in order to avoid over or under estimating the contribution from the water column.

Section 3.2.5.5 - Uncertainty Assessment:

The role of any uncertainty analysis should be to assess the error in the model calculations, which is important because the magnitude of the model needs to be considered when interpreting the results of the model calculations for management purposes. Instead of using Monte Carlo Simulations to describe the uncertainty in model variables, ~~The~~ comparisons of predicted and observed results ~~in an uncertainty analysis~~ can be plotted as a distribution of predicted and observed ratios. A rigorous analysis of observed versus predicted data variability may be used to develop estimates of model uncertainty. This avoids inherent problems in attempting a traditional uncertainty analysis for food web modeling where ~~In an uncertainty analysis,~~ many parameters are correlated, ~~, so it is~~making it difficult to maintain ~~apply Monte Carlo simulations of uncertainty~~ independence in model variables. ~~One of the key requirements is that the model state variables are independent.~~ Examples of related state variables are animal size, growth rate, lipid content and feeding rate. Parameters you can include and still maintain independence are limited and include lipid content, temperature and Kow. Evaluating differences between observed and predicted concentrations will also resolve the limitations of the current version of the food web model, which does not evaluate true distributions of data but instead evaluates only

the standard error around the calculation of mean values~~In addition, a true uncertainty analysis will evaluate uncertainty in each model parameter.~~

~~For uncertainty analysis, the parameters you can include and still maintain independence include lipid content, temperature and K_{ow}.~~ ~~He does not recommend~~ Conducting ~~an~~ uncertainty analysis on prey composition is not recommended , but ~~can use~~ a sensitivity analysis can be used to see how much of a difference prey assumptions would make (see above comment).

Section 3.5.2 – Application of the Model at Smaller Spatial Scales:

The harbor wide calibrated Portland Harbor food web model was executed at smaller scales and on laboratory test data. In general, the fits (SPAFs) were poorer, and these results were not surprising given that the model is highly calibrated to harbor wide average values.

Summary Comments

The Arnot and Gobas model has been used with modification to 21 internal parameters. The Portland Harbor model is high calibrated with field data from the site. The calculations with the model appear to have been done correctly. The regression equations for predicting residues in tissue based upon concentration of chemical in the sediment were done correctly. There tends to be a lot scatter in the data, and I wonder if the regressions have any predictive ability. Many of the predictions were done for mixtures, and if the composition of the mixture changes across the harbor, then predictions will be inaccurate and have high uncertainty. (This issue is not addressed in Appendix E.) This influences both food web predictions and BSAF calculations. The inverse dependence of BSAFs upon concentration of the chemical in sediment on an organic carbon basis indicates that bioavailability changes with increasing concentration in the sediment. This issue isn't addressed in the SWAC calculation while the field measured BSAFs incorporate these influences.